

**BLANK PAGE**

MULTI-DIMENSIONAL MAXIMIZATION AND  
DYNAMIC PROGRAMMING

By

Richard Bellman

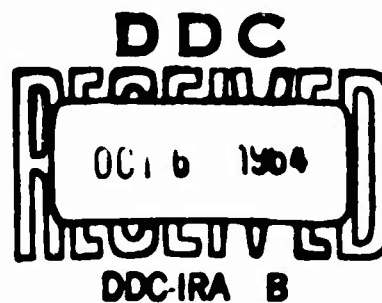
P-1086

*Boe* ✓  
May 14, 1957

Approved for OTS release

|            |     |      |   |
|------------|-----|------|---|
| COPY       | 1   | Of   | 1 |
| HARD COPY  | \$. | 1.00 |   |
| MICROFICHE | \$. | 0.50 |   |

*18p*



The RAND Corporation

1700 MAIN ST. • SANTA MONICA • CALIFORNIA

### SUMMARY

In this article we discuss some of the difficulties arising in multi-dimensional maximization problems and some of the special types of problems which can be treated by dynamic programming techniques. A brief discussion of linear programming is included.

## CONTENTS

|   |    |
|---|----|
| SUMMARY .....                                 | 11 |
| Section                                       |    |
| 1. INTRODUCTION .....                         | 1  |
| 2. ALLOCATION OF RESOURCES .....              | 6  |
| 3. SEARCH TECHNIQUES .....                    | 8  |
| 4. A PACKING PROBLEM .....                    | 9  |
| 5. A SMOOTHING PROBLEM .....                  | 9  |
| 6. AN EIGENVALUE PROBLEM .....                | 10 |
| 7. THE CALCULUS OF VARIATIONS .....           | 11 |
| 8. ALLOCATION OF TWO TYPES OF RESOURCES ..... | 11 |
| 9. LINEAR PROGRAMMING .....                   | 12 |
| 10. THE "FLOODING" TECHNIQUE .....            | 13 |
| REFERENCES .....                              | 15 |

## MULTI-DIMENSIONAL MAXIMIZATION AND DYNAMIC PROGRAMMING

By

Richard Bellman

### §1. INTRODUCTION

In the mathematical domain, the word "solution" does not have a unique meaning. As Poincaré put it, one generation's solutions are another generation's problems. In the first place, a problem may be considered to be solved if the existence of a solution can be demonstrated. In a number of fields, this step represents either an outstanding achievement, or a continued challenge. On the second level, a problem may be claimed to be solved if an algorithm exists for obtaining the solution. Most of the questions of analysis fall within this category. Finally, a problem may be considered solved if we possess a feasible algorithm for obtaining the solution. The word "feasible" is used here to describe a procedure which will yield the solution with desired accuracy at a reasonable cost in time.

This last is, of course, the only fully satisfying concept of a solution, and was so considered by Gauss. In more recent times, mathematicians have occasionally been a bit negligent in distinguishing between the second and third levels, although a great deal of energy has continually been directed towards the problem of raising problems from the first to the second level.

Fortunately, throughout the years, the incessant clamor of physicists, engineers, economists and others who have felt the pressure of producing numbers, has had the happy effect of

emphasizing to the mathematician the vast gap that exists between these last two levels of solution. The resultant challenge to the mathematician has created a tremendous resurgence of interest in a host of questions of theoretical and practical significance that arise when we attempt to proceed from the second to the third level. With the advent of modern computing devices, the scientists in all fields are afforded the opportunity to consider and resolve problems which formerly appeared as far distant in space and time as the star Sirius.

To illustrate the foregoing remarks, consider the familiar problem of solving a system of linear equations of the form

$$(1) \quad \sum_{j=1}^N a_{ij} x_j = c_i, \quad i=1,2,\dots,N.$$

It can readily be shown, on the basis of quite general and abstract theories that a solution of this system exists and is unique provided that the determinant of the system,  $|a_{ij}|$ , is non-zero. In addition, a number of properties of the solution, such as linear dependence upon the  $c_i$ , can be deduced. This is a first level solution.

A second level solution is based upon Cramer's rule which exhibits the solution as ratios of determinants. Thus

$$(2) \quad x_1 = \frac{\begin{vmatrix} c_1 & a_{12} & \cdots & a_{1N} \\ c_2 & a_{22} & \cdots & a_{2N} \\ \vdots & \vdots & & \vdots \\ c_N & a_{N2} & \cdots & a_{NN} \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1N} \\ a_{21} & a_{22} & \cdots & a_{2N} \\ \vdots & \vdots & & \vdots \\ a_{N1} & a_{N2} & \cdots & a_{NN} \end{vmatrix}}$$

with similar expressions for the other  $x_i$ .

This is an elegant representation of the solution which plays an important role in many investigations. It is an algorithm since we know how to evaluate determinants. It is, however, certainly not a feasible algorithm for large  $N$  of the order of 100 or more. To see this, recall that a determinant of degree  $N$  expanded according to the usual rules possesses  $N!$  terms. This proliferation of terms introduces two major difficulties.

To begin with, it consumes an appreciable amount of time to add  $100!$  numbers together. To give some crude idea of what a number like this means in computing time, observe that Stirling's formula states that

$$(3) \quad 100! \approx (100/e)^{100} \sqrt{200\pi}$$

which means that  $100! > (25)^{100} \approx 10^{140}$ . Consequently, if we had a super machine which could multiply 100 numbers together and add the result to another number in  $10^{-10}$  seconds, the evaluation of a  $100 \times 100$  determinant would consume  $10^{130}$  seconds. Convert this into minutes, hours, years, or millenia, and the result is still awe-inspiring.

Assuming that some magic device has been developed which permits  $100!$  operations in a short span of time, we are still faced by the ogre of "round-off error". Every time we multiply two numbers containing ten significant figures apiece together and round-off the answer to ten significant figures, we commit an error, and similarly every time we add two such numbers together we may have to round-off the answer. An error of order of magnitude  $10^{-10}$  committed  $100!$  times may eventually overwhelm

the answer, leaving us absolutely nothing to show for the time and effort expended.

It follows that the formula of Cramer, so valuable for theoretical analysis, is totally useless for computational purposes. We find ourselves then in the paradoxical situation where an explicit representation of the exact solution of a problem must be discarded, to be replaced by techniques for obtaining the approximate solution. The reader who is interested in finding out how the modern mathematician escapes from this dilemma will enjoy the excellent expository article by Forsythe, [1].

Having seen that the problem of the solution of linear systems of equations can exist simultaneously in all three modes of solution, let us now turn to another equally basic problem which is also tri-valent. We shall devote the remainder of this article to a discussion of this problem.

Consider the problem of determining the maximum value of a function of  $N$  variables  $F(x_1, x_2, \dots, x_N)$ , where the independent variables  $x_i$  are constrained to lie within some region  $R$  which may be defined by means of a set of inequalities of the form

$$(4) \quad G_i(x_1, x_2, \dots, x_N) \leq 0, \quad i = 1, 2, \dots, M.$$

As a first level problem, the question is readily resolved. If the function is continuous over the region  $R$ , and if  $R$  is bounded, a fundamental theorem of Weierstrass tells us that the maximum is actually assumed.

If each of the variables  $x_i$  assumes only a finite set of



values, then the existence of a maximum is rudimentary. Let us observe parenthetically that it is usually much more difficult to determine the maximum when the maximization is over a discrete set of points than it is when continuous variation is permitted; see, for example, Tompkins, [15].

Let us now add some further assumptions. Let  $P$  be a differentiable function of the variables involved, and suppose that the maximum does not occur on the boundary of the region  $R$ . Then the point at which the maximum occurs is to be found as one of the solutions of the system of simultaneous equations

$$(5) \quad \frac{\partial P}{\partial x_i} = 0, \quad i = 1, 2, \dots, N.$$

If  $P$  is a quadratic form, this system of equations is linear, which means that we possess feasible algorithms. If, as is generally the case, the system of equations is nonlinear, we are forced to use various iterative techniques, such as the gradient method, which are only occasionally successful; see the interesting article by Rosenbloom, [14].

If some of the maximizing values lie on boundaries, or if some of the variables assume only a discrete set of values, the methods of calculus are only partially operative. At the present time, most problems of these types appear to be hedged about with insurmountable obstacles.

In the remainder of the paper we shall discuss some particular multi-dimensional maximization problems which can be resolved by means of a judicious combination of high-speed

computers and the theory of dynamic programming, [1], [2]. We shall also briefly mention a quite important type of problem that can be treated by means of the theory of linear programming and the simplex technique of Dantzig, [9]. In conclusion, we shall touch upon the "flooding" technique of Boldyreff, [8].

## §2. ALLOCATION OF RESOURCES

Let us now consider the problem of maximizing the function

$$(1) \quad R_N = \sum_{i=1}^N g_i(x_i)$$

over all values of the  $x_i$  satisfying the inequalities

$$(2) \quad \begin{aligned} x_1 + x_2 + \dots + x_N &\leq c \\ x_i &\geq 0. \end{aligned}$$

The individual functions  $g_i(x_i)$  are assumed to be continuous for  $x_i$  in the interval  $[0, c]$ . This is the only condition we shall impose, since we shall not employ any of the methods of calculus.

This problem may be considered to arise in the following fashion. We have a certain quantity of resources which we wish to allocate among  $N$  activities, with a quantity  $x_i$  going to the  $i$ th activity. The return from this allocation is measured by the functions  $g_i(x_i)$ . The problem is to determine the  $x_i$  so as to maximize the total return.

Instead of the usual approach which views this as an isolated problem, with  $c$  a fixed quantity and  $N$  a fixed number, we consider the entire set of problems of this type, allowing  $N$  to assume any

integer value, and  $c$  any non-negative value. We then define the function of two variables,  $c$  and  $N$ , by the relation

$$(3) \quad f_N(c) = \text{Max}_{\{x_1\}} F(x_1, x_2, \dots, x_N),$$

where the maximum is taken over the region defined by (2),  $c \geq 0$  and  $N = 1, 2, \dots$ .

The simplest member of this family of functions is  $f_1(c)$ , determined by the relation

$$(4) \quad f_1(c) = \text{Max}_{0 \leq x_1 \leq c_1} g_1(x_1).$$

If we can find a relation connecting  $f_N(c)$  and  $f_{N-1}(c)$  we can consider ourselves to have a second level solution, since iteration of this relation will determine  $f_N(c)$  as a function of  $f_1(c)$  which is known.

To obtain this recurrence relation, let us reason in the following fashion. Assuming that we have allocated a certain quantity of resources,  $x_N$ , to the  $N$ th activity, we will have a quantity of resources  $x - x_N$  remaining which we wish to divide among the remaining activities. It is clear that the remaining allocations are to be made so as to maximize the return from the remaining  $N-1$  activities. Consequently, whatever the choice of  $x_N$  we must have the relation

$$(5) \quad f_N(x) = g_N(x_N) + f_{N-1}(x - x_N)$$

Since  $x_N$  is to be chosen so as to maximize the overall return, we must have the relation

$$(6) \quad f_N(x) = \max_{0 \leq x_N \leq x} \left[ g_N(x_N) + f_{N-1}(x-x_N) \right]$$

the desired recurrence relation.

The argument we have used is a particular case of the "principle of optimality", cf. [1], [2].

Since  $f_1(x)$  is determined from (4), we may compute  $f_2(x)$  from the above relation. Having determined  $f_2(x)$ , we determine  $f_3(x)$ , and so on. This determination is by means of a digital computer.

We have thus replaced the original N-dimensional maximization problems. Experience with a variety of problems has shown that this is a feasible algorithm and that we do indeed have a third level solution.

### §3. SEARCH TECHNIQUES

If the functions  $g_N(x_N)$  have no special properties, no better procedure for finding the maximum exists than the straightforward examination of a set of equally spaced points in the x-range of interest. If, however, it is known that the function  $g_N(x_N) + f_{N-1}(x-x_N)$  is concave then the problem of determining an optimal search pattern becomes meaningful. This problem, itself a dynamic programming problem, has been treated by S. Johnson and J. Kiefer in independent publications, [12], [13]. Interestingly enough, the Fibonacci numbers play a role in the solution. The corresponding problem concerning the location of the zero of a concave function has been treated by Gross and Johnson.

#### §4. A PACKING PROBLEM

A particularly interesting application of the method outlined in §2 is to the problem of determining the maximum of the linear function

$$(1) \quad R_N = \sum_{i=1}^N v_i x_i$$

subject to the constraints

$$(2) \quad \begin{aligned} x_i &= 0, 1, 2, \dots, \quad i = 1, 2, \dots, N, \\ \sum_{i=1}^N w_i x_i &\leq w. \end{aligned}$$

The problem is a good deal trickier than it appears due to the condition that the  $x_i$  assume only zero or integral values. Although no analytic solution has yet been given, the computational solution is readily obtained using functional equations.

The question arises in the determination of a most valuable cargo subject to weight restrictions.

#### §5. A SMOOTHING PROBLEM

As a further application of the method, consider the problem of determining the minimum of the function

$$(1) \quad \sum_{k=1}^N g_k(x_k - c_{k-1}) + \sum_{k=1}^N h_k(x_k)$$

over all values of the  $x_i$ . This is a particular case of a "smoothing" problem where a sequence of levels of activity must be determined in such a way as to minimize the total cost of maintaining these levels plus the cost of transition from one level to the succeeding.

Introducing the sequence of functions,  $f_N(c)$ , defined by the relations

$$(2) \quad f_N(c) = \min_{x_N} \left[ g_N(x_N - c) + h_N(x_N) \right],$$

$$f_k(c) = \min_{x_k} \left[ g_k(x_k - c) + h_k(x_k) + f_{k+1}(x_k) \right],$$

we readily obtain the computational solution.

#### §6. AN EIGENVALUE PROBLEM

The problem of determining values of  $\lambda$  which permit non-trivial solutions of the boundary-value problem

$$(1) \quad u'' + \lambda a(t)u = 0, \quad u(0) = u(1) = 0,$$

can, under simple assumptions concerning  $a(t)$ , be reduced to the problem of determining the relative minima of the functional

$$(2) \quad \int_0^1 u'^2 dt$$

subject to the constraints

$$(3) \quad \int_0^1 a(t)u^2 dt = 1, \quad u(0) = u(1) = 0.$$

A discrete version of this problem is the problem of determining the relative minimum of the quadratic form

$$(4) \quad \sum_{k=1}^N (x_k - x_{k-1})^2$$

subject to the constraint

$$(5) \quad \sum_{k=1}^{N-1} a_k x_k^2 = 1, \quad x_0 = 0, \quad x_N = 0.$$

The problem of determining the absolute minimum, which is to say the smallest characteristic value, is one which can be approached by means of the technique exhibited above. Further details may be found in [1], [3].

### §7. THE CALCULUS OF VARIATIONS

Similarly, more general problems in the calculus of variations, involving the minimization over all functions  $y$  of a functional of the form

$$(1) \quad \int_0^T F(x, y, s) ds$$

subject to constraints of the form

$$(2) \quad \frac{dx}{dt} = G(x, y, t), \quad x(0) = c,$$

and

$$(3) \quad R(x, y) \leq 0$$

can be treated by means of functional equation techniques. Since any discussion of this would take us too far afield, we refer the reader to [1], [4], [5], where further discussion and applications may be found.

### §8. ALLOCATION OF TWO TYPES OF RESOURCES

Let us now consider the problem of maximizing the function

$$(1) \quad R_N = \sum_{i=1}^N g_i(x_i, y_i),$$

subject to the constraints

$$(2) \quad \begin{aligned} \sum x_i &\leq x, & x_i &\geq 0, \\ \sum y_i &\leq y, & y_i &\geq 0, \end{aligned}$$

which arises from allocation problems involving two types of resources.

Introduce the sequence of functions of two variables

$$(3) \quad f_k(x, y) = \text{Max } R_k,$$

$$k = 1, 2, \dots, \quad x, y, \geq 0,$$

where the maximization is over the above region.

Then

$$(4) \quad f_k(x, y) = \text{Max}_{\substack{0 \leq x_k \leq x \\ 0 \leq y_k \leq y}} \left[ g_k(x_k, y_k) + f_{k-1}(x-x_k, y-y_k) \right]$$

with

$$(5) \quad f_1(x, y) = \text{Max}_{\substack{0 \leq x_1 \leq x \\ 0 \leq y_1 \leq y}} g_1(x_1, y_1).$$

Problems involving the computation of sequences of functions of two or more variables are very much more complicated than those involving the computational solution of sequences of functions of one variable. For problems of the type described above, a combination of the Lagrange multiplier technique and the functional equation method of dynamic programming has proved quite successful, see [6].

## §9. LINEAR PROGRAMMING

The problem of allocating a large number of resources leads to the question of maximizing the linear function

$$(1) \quad \sum_{i=1}^N b_i x_i$$



subject to a series of constraints of the form

$$(2) \quad \sum_{j=1}^N a_{1j} x_j \leq c_1, \quad 1 = 1, 2, \dots, M,$$

under suitable assumptions of linearity concerning the cost and return functions. In addition, a great variety of other problems that arise in economic and industrial life involving the scheduling of operations can be cast in this form.

Although the functional equation approach yields a second level solution, in the form of an algorithm based upon recurrence relations, this approach is not at all feasible for large values of  $M$ .

A method of an entirely different type has been developed by G. Dantzig, and considerably extended by Dantzig and others. This method, called the "simplex method", is based upon the linearity of the functions involved, see [9].

#### §10. THE "FLOODING" TECHNIQUE

A number of scheduling problems arising in transportation theory, see [7], give rise to problems of the type mentioned above. Some of these, such as the Hitchcock-Koopmans transportation problem may be solved in very elegant and rapid fashion by means of the simplex technique. Some, on the other hand, require iterative techniques specifically adapted to the problem at hand. A technique of this type, reminiscent of the relaxation technique of Southwell, has recently been developed by Boldyreff, the "flooding technique". A discussion of it may be found in [8].

Work on the same problem which gave rise to this technique,

P-1086  
5-14-57  
~~-14-~~

a transportation problem of Harris, has also resulted in the development of modifications of the simplex technique which are useful in treating a number of other problems as well. These results, due to Ford and Fulkerson, may be found in [10].

### REFERENCES

1. R. Bellman, Dynamic Programming, Princeton University Press, 1957.
2. ———, The Theory of Dynamic Programming, Bull. Amer. Math. Soc., Vol. 60, 1954, pp. 503-516.
3. ———, Eigenvalues and Functional Equations, Proc. Amer. Math. Soc., Vol. —, 1957, pp. 68-72.
4. ———, Dynamic Programming and Its Application to Variational Problems in Mathematical Economics, Symposium on the Calculus of Variations and Its Applications, Amer. Math. Soc., 1955, Chicago.
5. ———, On the Application of the Theory of Dynamic Programming to the Study of Control Processes, Symposium on Nonlinear Control Processes, Brooklyn Polytechnic Institute, 1956.
6. ———, Dynamic Programming and Lagrange Multipliers Proc. Nat. Acad. Sci., Vol. 42, 1956, pp. 767-769.
7. ———, Mathematical Aspects of Scheduling Theory, Jour. Soc. Ind. Applied Mathematics, Vol. 4, 1956, pp. 168-205.
8. A. Boldyreff, Determination of the Maximal Steady-State Flow of Traffic Through a Railway Network, Jour. Oper. Res. Soc. Amer., Vol. 3, 1953, pp. 443-465.
9. G. B. Dantzig, Application of the Simplex Method to a Transportation Problem, Chap. XII, pp. 359-373, Activity Analysis of Production and Allocation, Wiley, 1951.
10. L. Ford and R. Fulkerson, A Simple Algorithm for Finding Maximal Network Flow and Application to the Hitchcock Problem, Can. Jour. of Math., to appear.
11. G. E. Forsythe, Solving Linear Systems can be Interesting, Bull. Amer. Math. Soc., Vol. 59, 1953, pp. 299-329.
12. S. Johnson, Optimal Search is Fibonacciian, The RAND Corporation Paper P-856, May 4, 1956.
13. J. Kiefer, Sequential Minmax Search for a Maximum, Proc. Amer. Math. Soc., Vol. 4, 1953, pp. 502-506.
14. P. Rosenbloom, The Method of Steepest Descent, Chapter 11, Proceedings of the Sixth Symposium in Applied Mathematics, Amer. Math. Soc., McGraw-Hill, 1956.
15. C. B. Tompkins, Machine Attacks on Problems Whose Variables Are Permutations, *ibid*, Chapter 14.

**BLANK PAGE**